

ИССЛЕДОВАНИЕ СВЯЗНОСТИ ОФИЦИАЛЬНОГО НАУЧНО-ОБРАЗОВАТЕЛЬНОГО ВЕБА

Ю.В. Чуйко, А.В. Чирков

Рабочая группа по вебометрике Института прикладных математических исследований КарНЦ РАН [1] в течение нескольких лет проводит исследования Веба, основанные на анализе гиперссылок. В 2008-2010 годах исследования были поддержаны Российским фондом фундаментальных исследований (проект № 08-07-00023а).

Для сбора, хранения и обработки информации о гиперссылках разработана и постоянно совершенствуется информационная система для вебометрических исследований (ИС4ВИ), работающая под управлением веб-сервера Apache с интегрированным модулем PHP и СУБД MySQL [2,3].

Объектами исследования в данной работе являются целевые множества, состоящие из 349 официальных сайтов учреждений и организаций РАН и 56 сайтов классических университетов РФ. Множества сайтов и соединяющих их гиперссылок представляются для исследования в виде ориентированного веб-графа, где вершины соответствуют сайтам, а дуги – уникальным гиперссылкам между ними. В этом случае параллельные дуги можно заменить одной дугой. Компоненты сильной связности в таком графе представляют множества сайтов, связанных гиперссылками так, что по ним можно перейти с любого на любой сайт данного множества. Целью исследования является нахождение компонент связности веб-графов и их диаметров для целевых множеств. При этом множества рассматриваются как по отдельности, так и в объединении. Это позволит установить, насколько сильно связаны (или разделены) между собой исследуемые сообщества сайтов, и построить гипотезу о том, насколько развито сотрудничество между владельцами сайтов. Для автоматизации проведения исследования и визуализации результатов было разработано приложение на языке Java 2 SDK, позволяющее в интерактивном режиме выбирать структуру веб-графа из множеств сайтов, находить и отображать компоненты связности. Процедура нахождения компонент связности основана на модификации алгоритма Флойда [4] поиска путей в ориентированном графе. Для визуализации представления веб-графов использована Java-библиотека Jung2 2.0 (<http://jung.sourceforge.net>).

При исследовании целевых множеств обнаружилось, что в каждом из них выделяется по одной нетривиальной компоненте сильной связности, остальные компоненты имеют мощности (число вершин), равные 1. Для сообщества академических сайтов выделена компонента мощностью 180 и диаметром 8. Причем визуально по количеству соединений на графе наиболее выделяются сайты www.ras.ru, www.jscc.ru, www.uran.ru, www.sbras.ru, www.ict.nsc.ru. Для университетов также найдена единственная нетривиальная компонента мощностью 28 и диаметром 5. Что интересно, www.msu.ru в данную компоненту не входит. Она строится вокруг www.ns.ru, www.psu.ru, www.tsu.ru, www.yu.ru, www.dvgu.ru, www.isu.ru, www.ksu.ru. Неожиданным результатом стало то, что при объединении двух сообществ не были выделены две отдельные нетривиальные компоненты сильной связности. Здесь снова была найдена единственная нетривиальная компонента мощностью 220 и диаметром 8. Понятно, что тогда она включает в себя компоненты сильной связности обоих сообществ, дополнительно к ним присоединились еще 12 сайтов. Причем сюда уже вошел и www.msu.ru.

Какие выводы можно здесь сделать. Первый, это то, что Академия наук и университеты сильно связаны между собой взаимным сотрудничеством. Второй вывод касается сайта МГУ www.msu.ru, можно предположить, что данный университет более связан с научным сообществом, чем с университетским.

Данные исследования находятся в данный момент в стадии развития. В дальнейшем предполагается установить, какие именно узлы являются значимыми точками сочленения для найденных компонент сильной связности, т.е. удаление каких узлов приведет к распаду компонент сильной связности на две и более. Интересно будет установить, какие именно сайты исследуемых сообществ оказались такими точками, связавшими компоненты академических и университетских сайтов.

ЛИТЕРАТУРА:

1. *ВЕБОМЕТРИКА*. Институт прикладных математических исследований КарНЦ РАН. [Электронный ресурс] – Режим доступа: <http://webometrics.krc.karelia.ru>.
2. А.А Печников, Н.Б. Луговая, Ю.В. Чуйко, И.Э. Косинец "Разработка инструментов для вебометрических исследований гиперссылок научных сайтов" // Вычислительные технологии. 2009. Том 14. №5. С. 66-78.
3. Программа для ЭВМ "Информационная система для вебометрических исследований". Свидетельство о государственной регистрации программы для ЭВМ "Информационная система для вебометрических исследований" Федеральной службы по интеллектуальной собственности, патентам и товарным знакам РФ № 2010610941 от 29 января 2010 г.
4. Н. Кристофидес Теория графов: алгоритмический подход. - М.: Мир, 1978. - 429 с.